The application of toxicogenomics for (drinking) water quality assessment

BTO 2011.049
November 2011

KWR Watercycle Research Institute

BTO

KWR Watercycle Research Institute

The application of toxicogenomics for (drinking) water quality assessment

BTO 2011.049
November 2011

Postbus 1072 3430 BB Nieuwegein The Netherlands P +31 30 606 95 11 F +31 30 606 11 65E info@kwrwater.nl

Colofon

Title
The application of toxicogenomics for (drinking) water quality assessment

Project number
B111744

Research programme
Chemical Water Quality

Project manager
Merijn Schriks

Client
Board of Commissioners

Quality Assurance
Annemarie van Wezel

Author(s)
Merijn Schriks (KWR)
Ron van der Oost (Waternet)
Corine Houtman (Het Waterlaboratorium)

Sent to
This report has been distributed among BTO-participants and is publicly available.

Summary

Due to anthropogenic activities, freshwater systems worldwide are exposed to thousands of compounds. Monitoring of priority pollutants is important, however, monitoring of all individual compounds would be practically impossible. Moreover, the effects on human health remains mostly unknown since compound toxicity data is often absent. With the release of increasing amounts of new (emerging) chemicals into the environment, new monitoring strategies are required to assess the effects of (drinking water relevant) chemicals on human health. Therefore, sensitive *in-vitro* bioassays have been developed which focus on specific physiological effects such as endocrine disruption or mutagenicity. The advantages of bioassays are that they directly determine the effect of (an entire mixture of) compounds present in an environmental sample instead of identifying single compounds. Thus, these bioassays are able to detect known and unknown compounds as long as they trigger the bioassay response. However, a disadvantage of applying these assays may be that they each focus on a (relatively narrow selection of) specific physiological endpoint and that the human relevance is often obscure.
The evolution of new technologies and the recent advances in the knowledge on DNA sequences and organisation (i.e. genomics) have enabled the development of new holistic tools, such as DNA microarrays.

In this report, an overview is given of innovative 'omic' methods and the mechanisms behind certain genomics technologies are explained. The main focus is on the potential application of DNA microarrays that are able to measure the impact of toxic substances on gene expression, i.e. transcription of DNA to mRNA (transcriptomics). Practical information on this method, such as sensitivity, analysis time, responsiveness, specificity, etc. is discussed. An overview is presented of the potential applications of this technology with regard to water quality assessment. In addition, the limitations and challenges that need to be overcome are discussed.

The main conclusion of this report is that genomic approaches can help to understand the effects of emerging compounds on biota. At present, they are not intended to replace classical *in vitro* bioassays but can provide useful information e.g. in the mode of action (MOA) of compounds. It is recommended to establish consortia for (eco)toxicogenomics studies to overcome the challenges as mentioned in this report.

The following potential applications for genomics in water quality assessment are foreseen:

- Prediction of the toxicity of compounds that are newly detected in aquatic samples and of which possible toxic properties and MOA are unknown.
- Environmental monitoring of toxic pressure at locations relevant for the water sector, such as drinking water intake locations. For this purpose, the No Observed Transcriptional Level (NOTEL) concept is a helpful concept to mutually compare different sites.
- Pre-screening of chemicals or locations for relevant toxicological effects, which can be investigated in more depth with specific bioassays.

Contents

1 Introduction

Due to anthropogenic activities, freshwater systems worldwide are exposed to thousands of compounds. In the European Union, for example, there are more than 100,000 registered chemicals (EINECS list), of which 30,000-70,000 are in daily use (Schwarzenbach *et. al.*, 2006). About 300 million tons of synthetic compounds are used annually in industrial and consumer products, and partially find their way to natural waters (Schwarzenbach *et. al.*, 2006). Such contamination can become an increasing problem for aquatic ecology and e.g. drinking water production (Houtman, 2010, Schriks *et. al.*, 2010). Many of these compounds (such as hormones) raise concern, especially when the effects on various physiological endpoints are unknown. However, chemical analytical monitoring of all individual compounds would be practically impossible. Moreover, for the majority of the compounds the effects on biota remain unknown since toxicity data is often absent. Therefore, sensitive *in vitro* bioassays can be applied (e.g. van der Linden *et. al.*, 2008), which often focus on a specific physiological endpoint such as endocrine disruption (e.g. CALUX bioassays) or mutagenicity (e.g. Ames II test). The advantage of bioassays is that (i) they detect any unknown compound that triggers a specific biological effect and (ii) the effects of an entire mixture of compounds present in a sample can be determined. However, a disadvantage of applying these assays may be that they focus on a relatively narrow selection of physiological endpoints and that the ecological/human relevance often remains unclear.

The evolution of new technologies and the recent advances in the knowledge on DNA sequences and organisation (i.e. genomics) have enabled the development of new holistic tools, such as (DNA) microarrays. The term 'microarrays' is used for many different slides on which very small dots of a substance are spotted, but here the term alludes to slides with DNA spots, used for genomics. With microarrays, the expression level of genes (transcriptomics) can be determined, for example after exposure of cells or an organism to a chemical. Exposure to a toxic compound will result in a change in the transcription level (gene expression) of a cell. It has been claimed that microarrays offer the possibility of a very fast determination of subtle, primary effects of a chemical, occurring at low levels of exposure Burczynski, 2003). Moreover, validated expression profiles of exposure and effect (phenotypic anchoring) of individual chemicals in certain species may be very useful for risk assessment purposes. Since thousands of genes can be tested at once, a very broad spectrum of possible effects can be covered simultaneously. A microarray experiment could therefore theoretically substitute a number of different *in vitro* bioassays. Alternatively, a response in specific genes or gene clusters may lead to a tailor-fit development of a very specific *in vitro* bioassay for an endpoint currently not available yet. Microarrays therefore clearly have a theoretical potential to solve some of the problems with current techniques for monitoring toxicity.

The present report aims to provide
(i) a state-of-the-art overview of microarray techniques and
(ii) an evaluation if and how these techniques can be applied for (drinking) water quality assessment.

2 Technical background

2.1 Biology

Genes are sections in the DNA that regulate the various physiological functions in organisms, mostly by coding for the amino-acid sequence of proteins. There can be many different genes in one long strand of DNA. Humans have 46 of such DNA-strands, also called chromosomes, on which 20,000-25,000 protein-coding genes have been identified so far (HGSC, 2004). Not all these genes are active in all cells all the time; most of them are only active when stimulated (as opposed to constitutional genes). This is how skin cells can differentiate into blood cells, for example, and how women start lactating only after delivery. The activity of genes is called "gene expression". It involves the copying of the gene code from the DNA to messenger RNA (mRNA); this copying process is called transcription and the entire collection of different mRNAs present in a biological sample at any one time is called the "transcriptome". The code of the mRNA is then translated into a protein at the ribosomes with the help of transfer RNA (tRNA). Figure 1 gives a schematic representation of these processes.

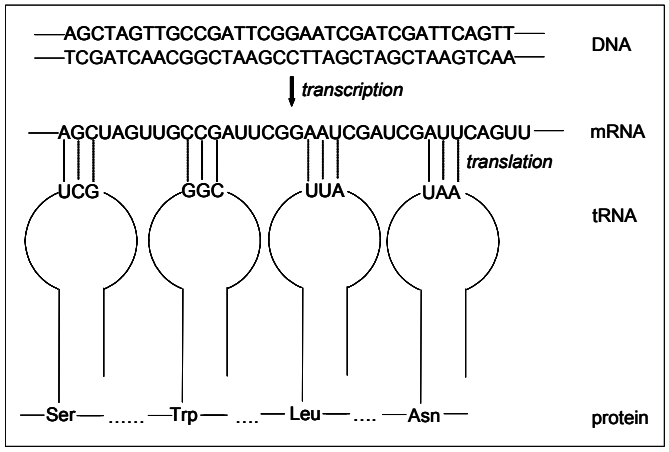

Figure 1. Schematic representation of the flow of genetic information, resulting in the production of a protein. DNA is double stranded, with the letters denoting the four bases forming the genetic code: adenine (A), guanine (G), cytosine (C) and thymine (T). One strand of the DNA is transcribed to messenger RNA (mRNA). RNA is single stranded and contains uracil (U) instead of thymine. A pairs with T or U, G pairs with C. Transfer RNA (tRNA) translates the genetic code of mRNA to an amino acid sequence, forming a protein. Examples of amino acids are shown here: serine (Ser), tryptophan (Trp), leucine (Leu), and asparagine (Asn). If a gene keeps on being stimulated, its DNA can be transcribed to mRNA multiple times, with more mRNA producing more of a certain protein. If the stimulation of a gene stops, no complementary mRNA is made anymore, and production of the related protein stops. Therefore, the expression of a gene is a measure of the stimulation it receives and the amount of mRNA present in a cell is a measure of gene expression.

2.2 Toxicology

If cells within an organism are exposed to a toxicant, many different reactions can take place, depending on the chemical and physical properties of the toxicant. Numerous biomarkers were developed as physiological indicators of exposure to specific toxicants. Biomarkers comprise a broad variety of endpoints ranging from biochemical responses to molecular alterations in protein or gene expression levels. In general, they are measured with quick, simple assays and provide biological evidence of exposure or effect (e.g. Van der Oost *et. al.*, 2003, Walker *et. al.*, 2006). One of the steps of these complex cascades of reactions is usually stimulation ("upregulation") or de-activation ("downregulation") of one or more genes. A well-known example of the involvement of genes in toxic action is the mechanism of action of estrogenic chemicals in male fish (e.g. Bogers *et. al.*, 2007). Estrogenic chemicals exert their action by binding to the cytosolic estrogen receptor. The resulting complex is translocated to the nucleus,

then binds to a specific promoter element on the DNA, resulting in activation of the expression of a gene next to this promoter, encoding for the production of the female egg yolk protein vitellogenin (VTG). This protein is normally produced only in female fish. The presence of this protein in male fish thus provides a measure (biomarker) for exposure to estrogenic substances.
Another well-investigated example is the activation of genes coding for the cytochrome P450 iso-enzymes that metabolize chemicals such as dioxins, PAHs and PCBs, when there are elevated levels of these compounds in the cell (e.g. Whyte *et. al.*, 2000). This is a protective mechanism, enabling the detoxification and elimination of chemicals even when cells are exposed to larger amounts than normal. A last example is the activation of genes coding for repair enzymes after the DNA has been damaged by chemicals or radiation.
Although the reliability of biomarkers and their relationship to population or community effects is sometimes questioned (Forbes *et. al.*, 2006), biomarkers can be important tools to be used alongside analytical chemical measurements and toxicity bioassays to strengthen monitoring programmes and risk assessment (Van der Oost *et. al.*, 2003). Each biomarker indicates the impact of a suite of chemical pollutants to one specific endpoint.
Genomic technologies, however, provide a more comprehensive picture of the toxic effects experienced by the organism since they target multiple molecular responses. The most important feature is that genomics methods are "holistic techniques" that do not investigate specific targets or endpoints but indicate the whole potential of adverse effects that may occur after exposure to toxic chemicals. Therefore, genomic technologies may represent a suite of biomarkers that -potentially- will be more informative, more specific and more sensitive than traditional toxicological endpoints. Genomic methods may be able to assess the potential risks of unknown toxic substances in the environment.

2.3 Genomics

Genomics technologies include three major fields at various levels of biological organisation (see also Table 1):

- *Gene expression or transcriptomics*

These technologies are able to assess differences in mRNA levels between two populations (exposed and reference). cDNA and oligonucleotide microarrays, serial analysis of gene expression (SAGE), differential display (DD) and, increasingly, high-throughput sequencing are used to compare steady-state levels of mRNA at a given time in the cell.

- *Protein expression or proteomics*

Changes in gene expression may lead to changes in protein levels, although it is also possible that protein levels are modulated independently of mRNA levels by time point and/or organism. Two-dimensional polyacrylamide gel electrophoresis (2D PAGE) or other chromatographic separation techniques such as liquid chromatography (LC) coupled to mass spectrometry (MS), can be utilized to identify proteins and quantify protein level changes.

- *Metabolite levels or metabolomics*

Many proteins are enzymes that catalyze certain cellular metabolic processes, which may lead to the formation of small-molecule metabolites. Metabolomics is the systematic study of these unique chemical fingerprints. Metabolites can be measured using various identification techniques such as LC-MS/MS or nuclear magnetic resonance (NMR). However, interpretation is difficult since a large number of metabolites have not been identified or referenced (Poynton *et. al.*, 2008).

Table 1. Techniques commonly used in genomics (based on Pyonton et. al., 2008 with modifications).

Technique	Reference
Transcriptomics	
cDNA and oligonucleotide microarray	Lockhart *et. al.*, 1996 and Schena *et. al.*, 1995
Differential display	Liang and Pardee, 1992
Serial analysis of gene expression (SAGE)	Yamamoto *et. al.*, 2001
Suppressive subtractive hybridization (SSH)	Diatchenko *et. al.*, 1996
Proteomics[a]	
2D gel electrophoresis (PAGE)	Gevaert and Vandekerckhove, 2000
DIGE	Unlu *et. al.*, 1997
Liquid chromatography	Abersold and Mann, 2003
Metabolomics[b]	
Nuclear Magnetic Resonance (NMR)	Nicholson *et. al.*, 1999
FT-ICR MS	Brown *et. al.*, 2005
ESI-MS	Plumb *et. al.*, 2002

[a]DIGE, 2D-Fluorescence Differential Gel Electrophoresis.
[b]FT-ICR MS, Fourier-Transform Ion Cyclotron Resonance Mass Spectrometry; ESI-MS, Electron Spray Ionisation Mass Spectrometry.

2.4 Transcriptomics with DNA microarrays

Methods

At present, most work is carried out –and thus most experience is available- in the field of transcriptome analysis (transcriptomics or gene expression analysis). Therefore, the present study will mainly focus on gene expression assays/transcriptomics (e.g. microarrays). Methods were already available to study gene expression (e.g. Real Time PCR), however these techniques are not designed for high throughput application (Burczynski, 2003). The greatest advantage of new hybridisation technologies (e.g. microarrays) is that they can be applied as holistic tools allowing simultaneous analysis of all genes of an organism. Hybridisation assays are slides of glass or plastic, on which tiny drops of DNA pieces (so-called probes) are spotted. After drying the slides, the pieces of DNA are immobilized on the surface. Every spot contains DNA of only one gene and it is known which spot contains which gene. The DNA pieces can either be complementary DNA (cDNA) fragments, or oligonucleotides. cDNA is a long strand of DNA (typically 500 to 5000 base pairs (Burczynski, 2003), representing a certain gene, made by copying the code of a piece of mRNA (i.e. reverse transcription). Oligonucleotides are short pieces of DNA, typically 20-60 bases long (Shioda, 2004), which code for only a very small, but representative, part of a gene. Hybridisation assays are available in three formats:

- macroarrays;
- microarrays;
- high-density oligonucleotide arrays (Vrana *et. al.*, 2003).

Microarrays in more detail

The term microarray is also used for the array technology in general, which can be confusing. Macroarrays are plastic slides or membranes, 8 by 12 cm, with cDNA of typically 200-8000 genes spotted (Vrana *et. al.*, 2003). Microarrays are usually glass microscope slides of 2.5 by 7.5 cm, on which cDNA of around 10,000 genes is spotted (Vrana *et. al.*, 2003). High-density oligonucleotide arrays, contain oligo-nucleotides that are synthesized directly on the coated surface (Figure 2). They can contain around 40,000 genes on a surface of 1 by 1 cm and can be acquired commercially from Affymetrix (e.g. GeneChips), Nimblegen or Agilent.

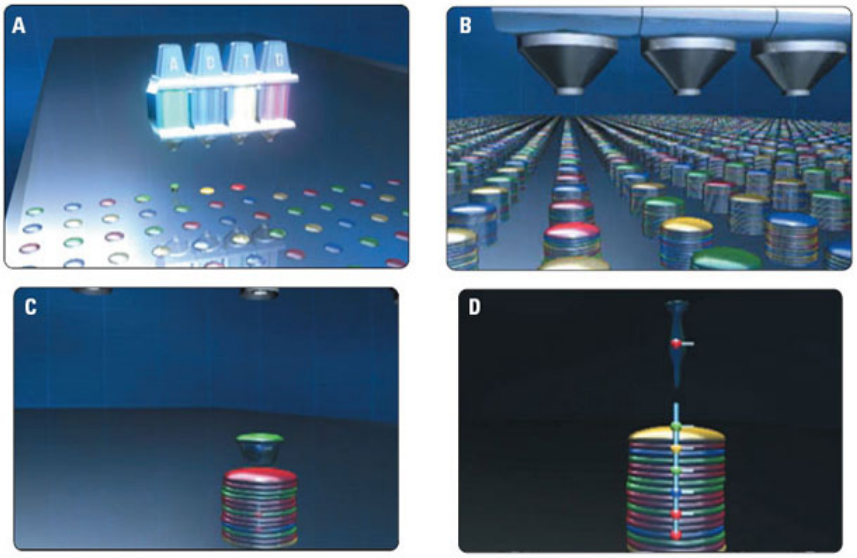

Figure 2. These four images communicate the general mechanism for oligo synthesis via inkjet printing. A: the first layer of nucleotides is deposited on the activated microarray surface. B: growth of the oligos is shown after multiple layers of nucleotides have been precisely printed. C: close-up of one oligo as a new base is being added to the chain, which is shown in figure D (courtesy of: Agilent website).

Because shorter sequences lead to fewer false matches, oligonucleotide arrays are more specific than cDNA arrays, and by containing around 20 different oligonucleotides for the same gene, sensitivity is increased as well (Affymetrix website). The arrays can be used to assess which genes in a tissue or cell culture are active and which are not. To this aim, present (m)RNA is isolated from the cells and single stranded cDNA is generated from this mRNA template by reverse transcription (Figure 3). The cDNA is radioactively labelled (macroarrays) or labelled with a fluorescent group (microarrays) (Vrana *et. al.*, 2003). The cDNA for oligonucleotide arrays is transcribed into cRNA, which is fragmented and labelled, either directly with fluorophores, or with biotin, which is conjugated to a fluorescent complex. This labelled cDNA or cRNA (the "target") is then spread over the array, making contact with all spots of immobilized cDNA or oligonucleotides (the "probes"). The labelled cDNA attaches to those pieces of DNA that are complementary ("hybridisation") and stay attached to the array during the following washing steps. Then, the array is analysed for fluorescent or radioactive spots, which show which genes were expressed in the cells.

With microarrays, a two-colour fluorescent labelling approach can be applied, to enable an efficient comparison of two samples on one and the same array. This approach avoids differences between different array slides, which can flaw the comparison of the different samples. Typically, different colours are used for control samples and treated samples, respectively. When both samples are hybridised to the same array, the different labelled cDNAs will compete for the immobilized DNA on the array. Relative differences in gene expression can be found by comparing differentially coloured spots.

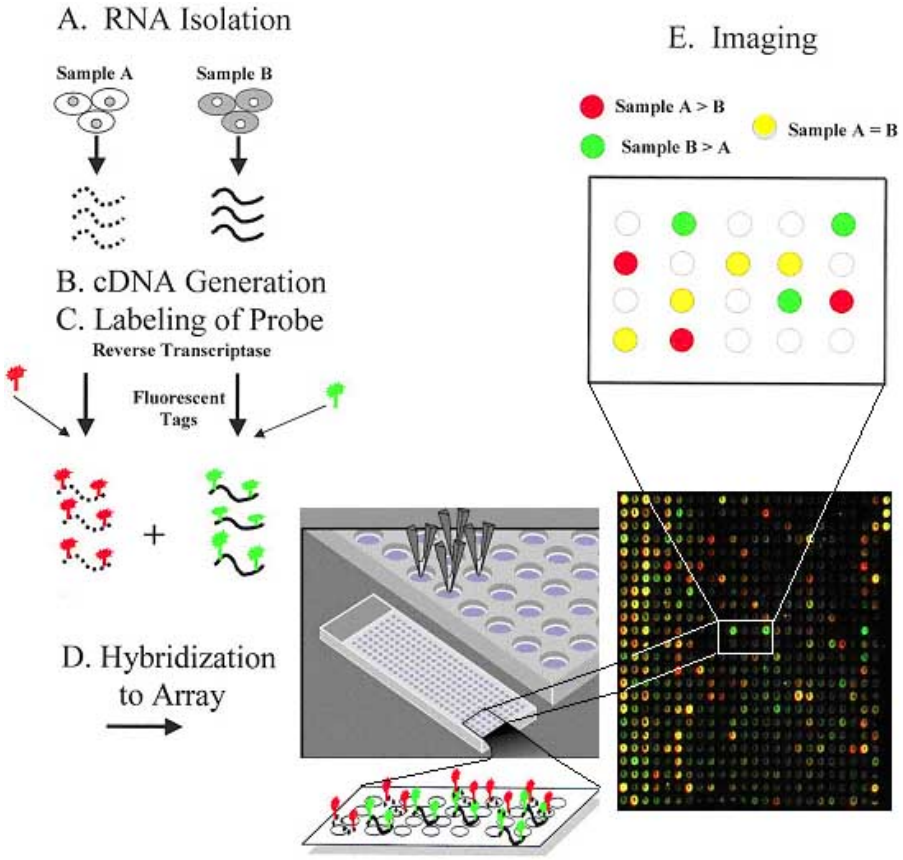

Figure 3. Schematic representation of the two-colour labelling approach on cDNA microarrays (http://www.fastol.com/~renkwitz/microarray_chips.

2.5 Sensitivity

A potential advantage of omics techniques is that they may be more sensitive than traditional physiological toxicity endpoints. Molecular, biochemical and cellular effects often occur at concentrations lower than whole organism effects such as mortality (Figure 4). This was demonstrated in a proteomic study using 2D-PAGE where a copper concentration-response (0-80 ppb, 24 hrs) was correlated to a measure of lyosomal damage in the gill tissue of the blue mussel (*Mytilus edulis)* (Shepard and Bradley, 2000). A protein expression pattern distinct from the control could be observed earlier at 20 and 40 ppb prior to physiological lysosomal damage. Results of the ILSI HESI Genotoxicity Working Group show that microarrays are equally or less sensitive than *in vitro* genotoxicity assays (Newton *et. al.*, 2004). Genotoxicity assays, such as the Ames II assay, are known to be relatively sensitive, with detection limits for certain genotoxic compounds below <500 ng/L; Heringa, *unpublished data*). However, this pattern may well be different for other toxicity endpoints.

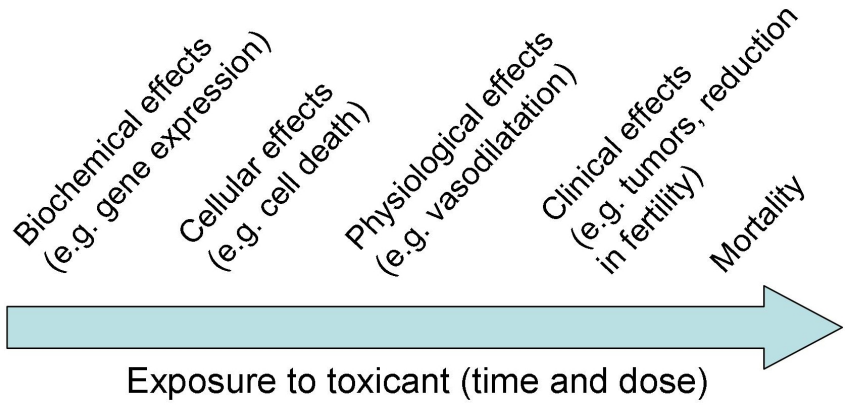

Figure 4. Cascade of events following toxicant exposure at a given time/dose.

2.6 Analysis time

The practical execution of a cell culture based microarray experiment generally takes about 3 days; cells need to be exposed for around 6-24 hours to create a change in mRNA levels; the isolation of mRNA, copying to cDNA, labelling of the cDNA and hybridization to the array takes about a day; initial analysis of the array data takes about another day (Newton *et. al.*, 2004). The analysis time also greatly depends on e.g. the amount of samples, the level of automation, detail of data processing and the number of endpoints to be studied. A microarray study generally requires less analysis time than traditional animal studies, but takes roughly the same time as *in vitro* assays such as the Ames test and CALUX bioassays. However, the follow-up bioinformatics analysis can be very time-consuming as long as standard protocols are lacking. In a DNA microarray, however, a very large number of endpoints can be analysed simultaneously, which could theoretically make the total analysis time of one microarray shorter than that of all related *in vitro* assays together.

2.7 Responsiveness

Poynton and co-workers (2008) suggest that genomics techniques may be used to predict chronic toxicity effects based on short-term exposure, because gene expression changes occur at an earlier stage or at a lower dose as compared to phenotypic responses. This is demonstrated in two studies in which water fleas (*Daphna magna*) were exposed to chemicals previously shown to disrupt moulting and reproduction, such as the fungicide fenarimol and zinc (Poynton *et. al.*, 2007; Soetaert *et. al.*, 2007). It was shown that after short term exposure (24 h or 4 days) a number of genes associated with the moulting process were down regulated. These exposures were much shorter than the classic 21 day Daphnia magna toxicity assay, suggesting that gene expression after short term exposure can be used to predict long-term effects. The early-warning responses of "omic" techniques, as well as the limitations in relevance for organisms and populations, are comparable to those of biomarkers.

2.8 Specificity

A limitation of the genomics techniques, as well as (*in vitro*) bioassays, is their inability to provide the identity of the responsible compounds. Specific bioassays and biomarkers are usually sensitive towards a certain group of compounds with a similar mechanism of toxicity. However, several studies have demonstrated that "omic" technologies have the capacity to differentiate between similar toxicants even within the same chemical class. An example is a study in which gene expression profiles of the biomarker vitellogenin (VTG) were compared to study endocrine disruptors. VTG induction is a good biomarker for estrogenic exposure in fish (Sumpter and Jobling, 1995), however, it does not discriminate between specific compounds. By applying a 132 gene cDNA microarray, Larkin *et. al.*, (2003a, 2003b) showed differential responses in gene expression in large mouth bass (*Micropterus salmoides*) exposed to the (xeno)estrogenic compounds 17β-estradiol, p,p'-DDE and 4-nonylphenol (4-NP). All three chemicals induced genes associated with VTG induction, however, 4-NP also induced other genes suggesting a secondary mode of action independent of activation of the estrogen receptor. This demonstrates that gene expression profiling can provide more specificity than the use of single biomarker, such as VTG.

2.9 Cross species extrapolation

Since it is possible that changes in expression of gene-sets or clusters among species lead to similar biochemical and/or physiological responses, genomics tools may be used to make interspecies comparisons. Fish, such as the zebrafish (*Danio rerio*) or the fathead minnow (*Pimephales promelas*), have been proposed as general models for toxicity assessment (Hill *et. al.*, 2005; Ankley *et. al.*, 2006). A correlation between toxicity symptoms described in mammals and gene expression profiles in the fathead minnow was observed by Wintz and co-workers (2006). Development of the Comparative Toxicogenomics Database (CTD; http://ctd.mdibl.org) facilitates interspecies extrapolation for risk assessment (Mattingly *et. al.*, 2006). Cross-species extrapolation based on a mechanistic understanding of toxicity achieved through genomics could allow invertebrate species to be used as alternatives for testing chemicals known to act through similar pathways across species. Studies that correlated toxicity values between invertebrate and vertebrate species provide reasonable reassurance that this may be possible (Guilhermino *et. al.*, 2000; Botsford, 2002). However, the data should be interpreted with caution, since there can be species- and time-specific responses.

3 Toxicogenomics

The incorporation of genomics tools and genomics knowledge in toxicology is known as toxicogenomics and it involves the study of toxicological processes at the gene level. In toxicogenomics, microarrays can be used to discover the mechanism of toxic action of chemicals (mechanistic toxicogenomics) or to predict the ultimate toxicity of chemicals (predictive toxicogenomics) (De Longueville *et. al.*, 2004).

Nuwaysir and co-workers (1999) define several areas at which toxicogenomics may be used including: (i) fundamental research to define modes of action, (ii) screening of novel emerging chemicals, (iii) environmental monitoring and (iv) risk assessment.
Apart from these, (v) the use of predictive toxicogenomics, i.e. microarrays as a toxicity screen, is the one of potential research fields.
For the development of such predictive toxicity screens, those genes have to be identified that characteristically change after exposure to a chemical. This is done by using arrays with as many genes as possible of the organism in question. The differences in gene expression between chemical-exposed and non-exposed cells are analyzed for different chemicals with a similar effect, to try and find a pattern in gene expression that is related to the toxic effect on tissue level. When a characteristic set of genes and their expression behaviour related to toxicity have been found, these specific genes can be spotted on arrays, to produce special toxicity-directed arrays. Such arrays would be easier and less expensive, as they contain fewer genes. By exposing cells to an unknown chemical and analyzing the expression in these cells of the specific "tox genes", toxic properties of the unknown chemical may be discovered. This approach, although valuable, does not exploit the full capabilities of omics' techniques, particularly their open nature, allowing discovery of unexpected changes (Van Aggelen *et. al.*, 2010).

4 Application of toxicogenomics

Toxicogenomics holds many promises for novel toxicity testing. As suggested by Poynton and co-workers (2008), combining traditional approaches with a genomic approach could provide a novel approach to predict the potential toxicity of a new chemical.
This chapter discusses research areas in which toxicogenomics tools could be or already are applied.

4.1 No Observable Transcriptional Effect Level (NOTEL) concept

For most toxic effects, chemicals may become relevant only at concentrations exceeding certain (environmental) concentration levels. The No Observable Transcriptional Effect Level (NOTEL) has been suggested as the concentration level of a chemical below which no significant changes in gene expression occur (Lobenhofer *et. al.*, 2004; Ankley *et. al.*, 2006). The NOTEL could provide a simple means to e.g. indicate a change in water quality. For example, if the expression level of certain genes reaches a defined threshold, an "alarm" could be triggered. A few studies have addressed the potential role of the NOTEL as an approach to identify the effects of low dose treatments and pinpointing a threshold dose below which no changes in gene expression are to be expected. A recent example of this approach is a study of Roling and co-workers (2007) demonstrated that a gradual drop in differentially expressed genes in mummichogs (*Fundulus heteroclitus*) at a chromium contaminated site undergoing remediation. NOTEL values for gene expression after cadmium exposure in various organisms were on average four times higher than the NOEC and eleven times lower than the LC_{50} values (Fedorenkova *et. al.*, 2010). A key issue when determining the NOTEL is the influence of confounding factors (see section "challenges").

4.2 Mode of action (MOA) approaches

Gene expression profiles can assist in the classification of chemicals based on their MOA. An interesting concept as suggested by Moggs (2005) is hierarchical clustering coupled with Gene Ontology (GO) to understand the categories of genes that are targeted by a toxicant and the physiological functions that might be impaired accordingly. GO is a systematic method to describe the molecular function, biological process and cellular components of each gene based on a structured classification system. Ontology data are stored in a central database. This public resource can be accessed via the GO consortium (www.geneontology.org). The investigator can use GO to determine which cellular or biological processes are affected by a stressor. The assignment of GO terms to a list of suggested differentially expressed genes can provide insight into biological processes that could be involved in a certain MOA. In a recent microarray study with fathead minnows (*Pimephales promela*) GO was used to predict the physiological effects caused by the estrogen 17β-estradiol (Larkin *et. al.*, 2007). The authors identified candidate targets of toxicity such as metabolism and regulation of cell growth. In addition, Van Boxtel and co-workers (2008) revealed a relevant mechanism of phenolic polybrominated diphenyl ether toxicity in Zebrafish by means of microarray analysis. However, for most species, functions of many genes are unknown and assignment of GO terms depends on annotated homologs in well-studied organisms which adds additional uncertainty (interspecies variability).

Another method to investigate the mode of action of a biological outcome of a chemical exposure is through phenotypic anchoring. This method is aimed at demonstrating that changes in gene expression levels can be associated with a toxicological/physiological syndrome. Without such a "link", a change in expression level of individual genes or gene clusters are of little or no toxicological/physiological relevance.

4.3 Environmental monitoring

Poynton *et. al.* (2008) suggested that genomic technologies can be applied to several different environmental monitoring settings such as an early warning system. The objective of general monitoring programmes is to provide an overview of water quality. Genomic technologies could contribute to these programmes, providing holistic tools to identify a decrease in water quality which could give rise to further detailed investigation such as chemical analysis. However, with the current research

emphasizing on technology development, the use of genomic technologies for environmental monitoring is still in its infancy.
A few studies have been carried out in the field outside controlled laboratory settings. An example is a study by Williams and co-workers (2003), focusing on the feasibility of toxicogenomics in a field situation. A 160 cDNA microarray was used to determine if gene expression in collected wild European flounder (*Platichthys flesus*) was different between a reference site and a site polluted with PAHs. Although high inter-individual variability was observed, the authors succeeded in identifying 11 genes that were differentially expressed in male flounders obtained from the polluted site. Similar results were obtained by Falciani and co-workers (2008) in the framework of the European GENIPOL project. Hepatic gene expression profiles of wild European flounder were used to predict the site of origin for the fish. The data as obtained implies that gene expression fingerprints are primarily attributable to variations in chemical pollutant response at the various sites, thus indicating potential use for environmental impact assessment. In another study, mummichogs (*Fundulus heteroclitus*) were used to investigate PAH resistance in a wild population inhabiting a site heavily polluted with PAHs. A 60 cDNA microarray was used to demonstrate that several biomarkers correlated strongly with PAH resistance (Meyer *et. al.*, 2005). In another study the effect of papermill effluent on gene expression in large mouth bass (*Micropterus salmoides*) was investigated. Female bass exhibited a downregulation of many reproductively important genes, which correlated with decreased reproduction and decreased hormone levels (Denslow *et. al., 2004*).
Thus, although gene expression is affected by many environmental factors, a subset of genes with altered expression can inform on stress responses. While these examples illustrate the feasibility of applying genomics in a field setting, a number of important factors need to be considered. As suggested by Poynton and co-workers (2008), in order to identify reliable marker genes relevant for the field situation, laboratory studies must first (i) provide information related to the time period that wild animals may be exposed and (ii) bridge the gap between genomic responses and environmental exposure (phenotype anchoring, see next section).

4.4 Risk assessment

The physiological/toxicological significance of these (genomic) biomarker responses must be clarified to allow efficient use in e.g. monitoring and regulatory decision making (Boverhof and Zacharewski, 2006) Potential applications of genomics for risk assessment purposes have been proposed (U.S. EPA, 2004), but to date genomic technologies have not been used in real-life risk assessment scenarios. Validated expression profiles of exposure and effect are required before they can be utilized in risk assessment. In other words, detailed concentration-response and time-course information is required to link long-term (adverse) effects with genomic profiles. This process of phenotypic anchoring of genomic response is crucial before genomics can be used in risk assessment (Paules, 2003).
REACH (**R**egistration, **E**valuation, **A**uthorisation and Restriction of **Ch**emical substances) is a new European Community Regulation on chemicals and their safe use (Regulation (EC) No 1907/2006). The aim of REACH is to improve the protection of human health and the environment through the better and earlier identification of the intrinsic properties of chemical substances. Therefore, toxic properties of a very large number of chemicals have to be analysed. A large number of research projects are focussed on the application of *in vitro* genomics techniques for an initial risk assessment of chemicals, in order to reduce the number of test animals that have to be sacrificed. The U.S. Environmental Protection Agency (U.S. EPA, 2002) currently accepts toxicogenomics data as part of a weight-of-evidence approach for establishing mechanisms of toxicity for regulated substances.

4.5 Investigation of biological processes controlled by gene networks

Although genome-wide RNA expression analysis is becoming a routine tool in biomedical research, extracting biological insight from such information remains a major challenge. Traditional strategies for gene expression analysis have focused on identifying individual genes that exhibit differences between two states of interest. Although useful, they fail to detect biological processes, such as metabolic pathways, transcriptional programs, and stress responses, that are distributed across an entire network of genes and subtle at the level of individual genes.
A powerful analytical method called Gene Set Enrichment Analysis (GSEA) has been developed for interpreting gene expression data based upon these networks (Subramanian *et. al.*, 2005). The method

derives its power by focusing on gene sets, that is, groups of genes that share common biological function, chromosomal location, or regulation (see Figure 5).

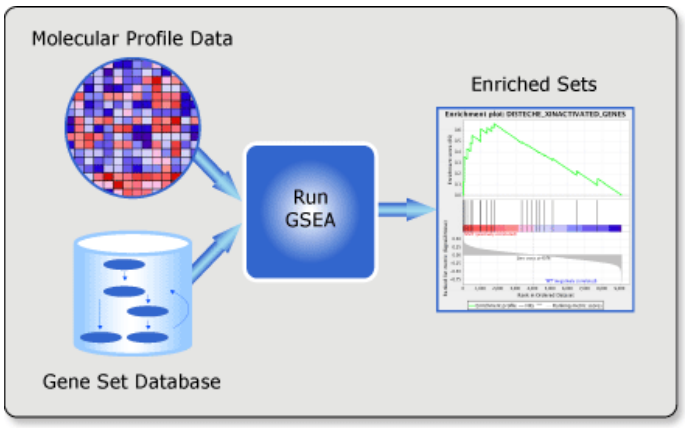

Figure 5: a schematic overview illustrating the GSEA method (source: GSEA website http://www.broadinstitute.org/gsea/).

As an example, GSEA yields insights into several cancer-related data sets, including leukaemia and lung cancer. Notably, where single-gene analysis finds little similarity between two independent studies of patient survival in lung cancer, GSEA reveals many biological pathways in common. The GSEA software is freely available on line (http://www.broadinstitute.org/gsea/), together with an initial database of 1,325 biologically defined gene sets.

There is an ongoing trend to study the responses on expression of (sub)networks of genes instead of individual genes. While simple associations between gene expression and outcome have been used successfully for biomarker identification, the interference is limited for each stressor. As a result, a separate interference is required for each stressor and there are dangers of overfitting a predictive model. Adverse Outcome Pathways (AOPs) provide a more generalized approach that can be used for multiple stressors, and thus may be more relevant for analysis of exposure of mixtures of chemicals in the environment.
A relevant example of AOP application has recently been described by Perkins and co-workers (2010). The authors utilized high dimensional omics data (gene, protein, metabolic, signalling) to characterize AOPs for flutamide, which is a chemical that disrupts the hormone balance, in fathead minnow (Figure 6). The identification of AOPs can thus assist in the development of targeted predictive screens for toxicity. The resulting networks have the potential to predict adverse effects to both the individual and to populations for a particular stressor or set of stressors. These outcome/endpoint predictions can be used to enhance the use of mode of action information in risk assessment, provide additional information to expand models of ecosystem effects, inform environmental protection abatement strategies, and support decisions about water quality criteria for a particular stressor or other chemicals with similar or related MOA or pathways or adverse effect(s). The impact of any stressor, or mixture of stressors, could be predicted by its effect on identified AOPs. It is important to recognize, however, that not all perturbations of an identified AOP will invariably lead to an adverse outcome.

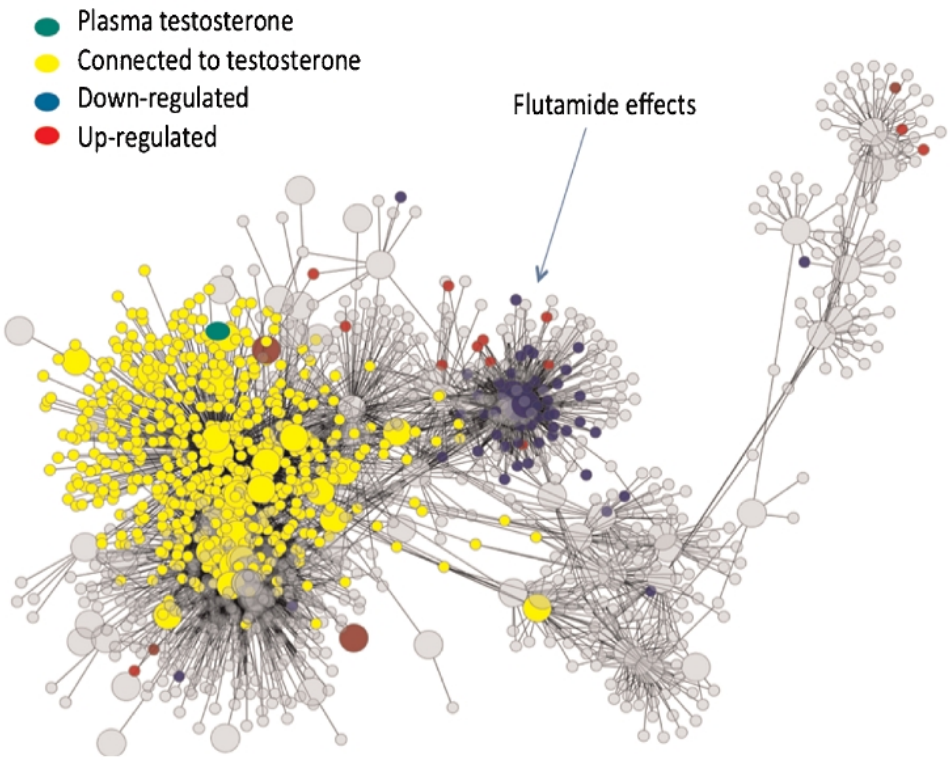

Figure 6: an example of an Adverse Outcome Pathway (AOP); Mapping the transcriptional response to flutamide exposure in fathead minnow (Perkins *et. al.*, 2011).

5 Challenges for toxicogenomics

The previous chapters presented the current possibilities and applications of genomics techniques. The discussed examples also demonstrate that there are challenges to overcome. The main challenges in the application of genomics were recently covered by Poynton and co-workers (2008) and are briefly described below.

5.1 Model system complexity

Although the mode of action (MOA) of specific chemicals may be similar between various species, their sensitivity may be very different. One approach to deal with this issue is to utilize a panel of different species. Some regulatory agencies such as the U.S. EPA have adopted this approach as part of their risk assessment process. Moreover, in the areas of both human and (eco)toxicological genomics, interpretation of the results can be hampered since not all genes are expressed in all tissues and organs. For example, muscle cells do not have any estrogen receptors, and will therefore not show any typical gene expression when exposed to estrogens. Muscle cells are therefore not suitable for a microarray tox-screen meant to detect estrogenicity in samples. As non-genotoxic carcinogens act through various mechanisms, which may not all be expressed in one cell-type, multiple cell types might be necessary to detect all non-genotoxic carcinogens in a sample.

5.2 Confounding factors

A major problem with genomics techniques is that genomic profiles can be indiscriminative and reveal both effects that are related to exposure to a potential toxicant as well as the effect of external confounding factors (Poynton *et. al*., 2008). This can be especially problematic when analyzing field organisms or cultured cells exposed to (environmental) water samples. Confounding factors include nutrient levels, acidity, temperature, salinity, dissolved organic carbon (DOC) and water hardness, which may modulate gene expression in an organism. Subtle differences in the surroundings of the cells, such as composition of culture media, can already result in gene expression changes (Shoida *et. al*., 2004), providing a source of misinterpretation of results. In order to discriminate between confounding factors and toxicant induced gene expression, it is necessary to understand the effects of (external) variables and how they alter gene expression. Well designed control experiments could aid to compensate for the effects of confounding factors. A number of studies explored how confounding factors themselves have distinct gene expression patterns. An example is a recent study of the effect of hypothermia on carp. An adaptive response to hypothermia, was reflected by robust changes in myriad gene expression changes (Chang *et. al*, 2005). By determining such stressor-specific expression profiles (or signatures), the use of data-reduction techniques (such as principle component analysis) may allow differentiation between gene expression changes induced by toxicants and those induced by confounding factors.

Several research groups have shown that expression profile signatures can be linked to known toxic mechanisms for tissues of treated animals, so-called phenotypic anchoring (e.g. Guerreiro *et. al*., 2003; Ellinger-Ziegelbauer *et. al*., 2004). This is most clearly demonstrated in studies in which animals are exposed to single compounds in well defined doses.

Under field conditions however, organisms are not exposed to a single compound but to a wide range of compounds, that can have a combined effect on the gene expression profiles. The major question is whether expression profiles of multiple compounds will reflect the combined expression profiles of single compounds and whether the toxicity responses can still be distinguished from the multitude of changed profiles. In an illustrative study by Krasnov and co-workers (2007) trout were exposed to Cd^{2+}, CCl_4 and pyrene. It was shown that at low level exposure, the expression profiles were additive and that Cd^{2+} and pyrene expression profiles could be distinguished from the mixture profile. However, it is likely that different combinations of chemicals (acting by different mechanisms and/or interacting with different endpoints) may result in different (or additive) effects on gene expression profiles in an organism. As suggested by Poynton and co-workers (2008) more studies are needed in order to understand the effects of mixtures on gene expression. It is unlikely that there will be a single pattern that explains the behaviour of all mixtures. Steinberg and co-workers (2008) claim that some more practical advantages of genomics techniques seem to emerge, since the toxicological assessment of

complex mixtures, such as effluents or sediments, as well as drugs seems feasible, especially when classical ecotoxicological tests have failed.

5.3 Mixtures

Several research groups have shown that expression profile signatures can be linked to known toxic mechanisms for tissues of treated animals, so-called phenotypic anchoring (e.g. Guerreiro *et. al.*, 2003; Ellinger-Ziegelbauer *et. al.*, 2004). This is most clearly demonstrated in studies in which animals are exposed to single compounds in well defined doses.

Under field conditions however, organisms are not exposed to a single compound but to a wide range of compounds, that can have a combined effect on the gene expression profiles. The major question is whether expression profiles of multiple compounds will reflect the combined expression profiles of single compounds and whether the toxicity responses can still be distinguished from the multitude of changed profiles. In an illustrative study by Krasnov and co-workers (2007) trout were exposed to Cd^{2+}, CCl_4 and pyrene. It was shown that at low level exposure, the expression profiles were additive and that Cd^{2+} and pyrene expression profiles could be distinguished from the mixture profile. However, it is likely that different combinations of chemicals (acting by different mechanisms and/or interacting with different endpoints) may result in different (or additive) effects on gene expression profiles in an organism. As suggested by Poynton and co-workers (2008) more studies are needed in order to understand the effects of mixtures on gene expression. It is unlikely that there will be a single pattern that explains the behaviour of all mixtures. Steinberg and co-workers (2008) claim that some more practical advantages of genomics techniques seem to emerge, since the toxicological assessment of complex mixtures, such as effluents or sediments, as well as drugs seems feasible, especially when classical ecotoxicological tests have failed.

5.4 Limited sequence data

An obstacle to the application of microarrays is the lack of genomic or cDNA sequence data for the organism of interest (Snape *et. al.*, 2004). For example, in ecotoxicology, investigators have used alternatives for constructing microarrays with limited sequence information (Snell *et. al.*, 2003). One method is to create cDNA libraries with selected transcripts that respond to (toxic) stress (Bultelle *et. al.*, 2002). These libraries focus on "relevant" transcripts and thus decrease the amount of sequencing needed. Another method, which has been explored in detail by Poynton and Vulpe (2009), is the use of so-called anonymous microarrays. Microarrays are printed with unsequenced cDNAs from a normalized library. Microarrays are carried out with only the transcripts that are responsive to a treatment, thus greatly decreasing the total number of transcripts. With the availability of high-throughput DNA sequencing technology, this limitation can be relatively easily bypassed through sequencing cDNA libraries derived from the organism of interest. Such comprehensive Expressed Sequence Tag (EST) data allows the design of, for example, oligonucleotide microarray probes, without necessarily requiring a full genome sequence.

5.5 Bioinformatics

The analysis of data derived from gene expression microarrays can be complex and remains an area of active research. A key first step is to identify differentially expressed genes (e.g. a polluted site versus a reference site). A lack in standardization of approaches to process bioinformation has led to difficulties when comparing results between laboratories. As originally suggested by Nuwasir and co-workers (1999), genes identified as differentially expressed can represent a "fingerprint" for specific exposure conditions. Because different laboratories use different data analysis methods, different experimental procedures, different apparatus, etc., a large variation may exist between data from different laboratories. Therefore, combination of data from different laboratories is necessary to build a large enough knowledge base for the development of applications, such as predictive toxicogenomics. Solid relationships between gene expression and toxicity can only be found when there are many comparable data on these endpoints. Therefore, a large effort is currently made, to develop a common format for data analysis (Newton *et. al.*, 2004). Already, standards have been introduced for the archiving of data, e.g. the MIAME (Minimum Information about a Microarray Experiment) standards for content and the MAGE (MicroArray and Gene Expression) data format standards (Freeman *et. al.*, 2004). Even though these standards are very useful for research purposes, they do not attempt to come near the quality criteria needed for a quantitative analytical measurement. The development of integrative software tools,

such as GSEA (see chapter 4.5), that analyse the impact of toxic stress on networks of genes involved in certain cellular processes, will be vital for "translation" of complex datasets into relevant information for regulators.
Ultimately, the linkage of genomic responses to relevant endpoints (e.g. behaviour, reproduction and development) has to be achieved. This will require better means of identification of an AOP, and also a better understanding of the relationship between the activation of combined AOPs and of the critical threshold levels of activation. The relative importance of short-term AOP activation and chronic activation in relation to toxicity is key, and thus temporal changes leading up to toxicity need to be understood.

5.6 Costs

Until recently, oligonucleotide microarrays were only commercially available for a small subset of species. Although some laboratories constructed their own oligo sets and printed them in-house for their specific species of interest, for most research groups this is financially not feasible. However, the development of high-throughput pyrosequencing, such as that available from 454 Life Sciences (Margulies *et. al.*, 2005; www.454.com), enables researchers to simultaneously obtain sequence information from a cDNA library sufficient to generate oligo probes representative of the library (Gowda *et. al.*, 2006). Recently, commercial suppliers such as Agilent and Nimblegen started offering custom oligonucleotide microarrays at reasonable costs. The feasibility of pyrosequencing to develop and utilize oligo microarrays has been recently demonstrated in a study of Denslow *et. al.* 2004.

6 Research needs for regulatory implementation

A number of factors currently limit widespread acceptance of genomics for regulatory applications. In addition to complex relationships between 'omic' responses and toxicological outcome, standardized, validated exposure assay and analysis procedures are lacking (Ankley *et. al.*, 2006). Consortia to establish standards for (eco)toxicogenomics studies need to overcome the challenges in this field and should help facilitate the integration of genomics into risk assessment and regulation (Poynton *et. al.*, 2008). Ankley and co-workers (2006) stated that a successful incorporation of toxicogenomics into regulatory frameworks may someday be regarded as the most important intellectual and practical contribution from this generation of (eco)toxicologists.

Workshops organized by various consortia of scientists and regulators identified both short- and long-term research needs for the implementation of the 'omic' technology in monitoring and risk assessment (Van Aggelen *et. al.*, 2010). The most important short-term needs are (i) formal standardization and validation of data collection, analysis, and presentation for standard test species, and (ii) generation of libraries of gene expression, proteomic, or metabolite profiling data based on a set of reference chemicals with well-defined, relevant MOAs. There have been important advances in recent years in the context of both of these fields. The long-term needs are (i) generation of genome sequence data for ecologically relevant species; and (ii) linkage of molecular and biochemical responses to adverse alterations in survival, growth and development, and reproduction.

Significant advances have been made in obtaining data for the development of reference gene expression profiling databases from species commonly used for regulatory assessments (Ankley *et. al.*, 2008), although much work remains. Because toxicogenomics data will be most valuable for predictive toxicology and elucidating toxicologically relevant MOAs for emerging chemicals, a future database should be based on toxicity testing and monitoring protocols commonly used for regulatory purposes (e.g., global pesticide registrations), as well as chemicals with well-known MOAs such as 17β-estradiol and dioxin (Van Aggelen *et. al.*, 2010). The availability of such a database will greatly enhance the possibilities for application of genomics for regulatory purposes.

7 Discussion on toxicogenomics for water quality monitoring

With the release of an increasing amount of new (emerging) chemicals into the environment, new methodology and risk assessment approaches are needed to assess the presence of (drinking water relevant) chemicals in the aquatic environment and their effects on human health (Houtman, 2010; Schriks *et. al.*, 2010). Genomic approaches might aid as new holistic tools to help understand the effects of such compounds and to monitor their occurrence in the aquatic environment. 'Omic' tools and associated endpoints are already improving the understanding of how individual chemicals and mixtures affect organisms (Van Aggelen *et. al.*, 2010). Ideally, omics data would reflect both the MOA and the subsequent deleterious outcome(s). To achieve this, the cascade of pathways associated with toxicity must be further defined, from a molecular initiating event (e.g., receptor binding) through subsequent biological alterations (reflected by 'omic' and cellular changes) that culminate in a deleterious outcome (NRC, 2007). As opposed to the initial 'hype' a more realistic understanding of the potential contribution of 'omics' to toxicology seems thus to be formed these days (NRC, 2007).
This chapter summarizes the possibilities for implementation of genomics for water quality assessment purposes.

7.1 Mechanisms of toxicity and predictive toxicology

Various well designed proof-of-principle studies as mentioned in this report have shown that gene expression profiling is able to suggest potential MOAs and predict exposure effects.
Effects considered most relevant for drinking water quality are (i) hormone disruption and reproductive toxicity (ii) genotoxicity and (iii) non-genotoxic carcinogenicity. Currently, a number of successful applications of toxicogenomics for these endpoints are published.

Endocrine disruption: A recent study investigated the effects of ethinylestradiol (EE2) and the anti-androgen flutamide on fathead minnow (Filby *et. al.*, 2007). Gene expression studies revealed clear differences in the affected genes and the authors suggested that microarray tools could be used to distinguish exposure to these chemicals in the environment. Another study demonstrated the ability to distinguish between 4 compounds (17β-estradiol, 4-nonylphenol, bisphenol A, ethinylestradiol) with estrogenic activity using common carp (*Cyprinus carpio*) (Moens *et. al.*, 2006).

Genotoxicity and non-genotoxic carcinogenicity: Expression patterns of different genotoxicants have been shown to be similar to a certain degree and potential genotoxicity may in fact be well evaluated with expression profiling (Kim *et. al.*, 2005). The prediction of non-genotoxic carcinogenicity seems to be limited at present, most likely related to the diversity in the mechanism of carcinogenesis (Kim *et. al.*, 2005). At present, differences have been found between gene expression patterns caused by genotoxicants and non-genotoxic carcinogens (Van Delft *et. al.*, 2004; Kim *et. al.*, 2005). Distinctions between these two types of carcinogens can therefore be made with this technology. Furthermore, data indicate that classes of genotoxic compounds -which act via different modes of action-, may be distinguished through gene expression (Newton *et. al.*, 2004). Likewise, chemical class-specific profiles have been found for liver toxicants (Hamadeh *et. al.* 2002a/b). More recently, Muellner and co-workers showed a number of altered transcript profiles in bromoacetic acid exposed human small intestine epithelial cells (Muellner *et. al.*, 2010). The majority of the altered genes are involved in DNA repair and cell cycle regulation, thus implicating that double stand DNA breaks are a feature of bromoacetic acid exposure. These studies show that omic tools are successfully used to investigate the action of known compounds with toxicities that are relevant for water quality. A first realistic application of 'omics' techniques could therefore be their use to predict the toxicity of compounds that are newly detected in aquatic samples and of which possible toxic properties and MOA are unknown.

7.2 Environmental monitoring

Application of omic tools for environmental monitoring requires higher discriminating power than needed for single compound studies, due to the complexity of environmental samples. At present, the sufficiency of the discriminating power of genomics techniques is still questionable.

There are, however, reports on the application of microarrays for studying complex mixtures and broad range of toxicities. For example, the NIEHS (National Institute of Environmental Health Sciences; USA) has developed the ToxChip: a cDNA microarray with 2,100 genes that are already known to exhibit changes resulting from exposure of cells (e.g. MCF-7 cells) to well-known toxicants (e.g. oxidant stressors, dioxins and environmental estrogens) (Medlin *et. al.*, 1999). Additionally, Kim and co-workers (2004) applied a cDNA microarray with yeast genes to evaluate the toxicity of water containing ash, which is a mixture of poorly defined compounds. Responses were found which were indicative of exposure to ROS (reactive oxygen species), metals, and genotoxicants. However, the results should be interpreted with caution, since the authors only based their conclusions on known functions of differentially expressed genes. For a more reliable assessment, established relations between gene expression and toxic effects should also be taken into account.

Furthermore, as discussed in section 4.3, there are examples of field studies in which discriminating power of used microarrays was sufficient to distinguish locations based on their level of contamination and subsequent effect on gene expression (Williams *et. al.*, 2003). These indicate that especially in case of specific biomarkers related to local pollution, genomics definitely have potential for environmental monitoring and water quality assessment.

7.3 NOTEL

The NOTEL concept (section 4.1) is defined as concentration level below which no significant changes in gene expression are observed. Until now, the concept has only been applied in studies with exposure to single compounds to identify their effects at low doses and to pinpoint threshold doses. For the purpose of water quality assessment, the concept might also be a helpful means to investigate the impact of environmental mixtures.

In the last decades, several aspects of the quality of surface waters and drinking water sources have improved considerably. Classical contaminants and high concentrations of severely toxic pesticides and spills of industrial contaminants occur less often. The focus of water quality control has shifted towards the investigation and monitoring of diffuse contamination and low concentrations of biologically active compounds, potentially with, alone or as mixture, subtle biological effects. The NOTEL concept might be applied to assess the combined impact of complex mixtures of a multitude of compounds at low concentrations simultaneously present in water samples. In such an approach, gene-expression changes induced by water samples and reference samples (blanks) (or extracts thereof) are tested on microarrays. The threshold needed to change gene expression is expressed as concentration factor of the sample (instead of concentration of compound, as done in studies with single compounds). Difference in concentration factors, between environmental water samples and blanks, indicates the impact (toxic pressure) of the combined mixture of contaminants in the water samples on gene expression. In this way, the NOTEL concept may provide a generic impression of the toxic pressure of contamination in investigated samples or at locations.

7.4 Genomics as a tool for monitoring in relation to *in vitro* bioassays

As freshwater systems are contaminated with thousands of different chemicals, chemical monitoring of all of them is not feasible. In the last years, sensitive and specific *in vitro* bioassays have been implemented more and more to monitor the presence of compounds with the most relevant toxic potencies.

Since thousands of genes can be tested at once, microarrays have sometimes been mentioned as candidates to replace bioassay measurements. Large batteries of *in vitro* bioassays could then theoretically be replaced by one microarray, saving costs and improving usefulness for risk assessment. This study has not found indications that such expectations are realistic. *In vitro* bioassays indeed focus only on a relatively narrow selection of physiological endpoints. However, this enables researchers to specifically study those effects that are already known to be relevant, in optimised (cell) systems, whereas such effects might remain indiscriminate in genomics approaches. In addition, *in vitro* bioassays offer the possibility to study effects in a quantitative manner, where microarrays give insights on a more qualitative level.

At present, a realistic application of 'omics' techniques may be their use in pre-screening chemicals and mixtures for prioritization in further tests (Ankley *et. al.* 2006). In addition, a microarray study to survey *locations* for (groups of) genes influenced may then be used as validation for the choice for specific *in vitro* bioassays relevant for the ecosystem of study.

7.5 Conclusions and outlook

Table 2 provides a summary of the possibilities explored in this chapter and a comparison of the possibilities of current water quality monitoring approaches (chemical analysis and bioassays) and genomics tools. Chemical analysis remains the best option for the identification of target chemical compounds. It does however not give any information on effects and unknown compounds. Bioassays, on the other hand, do not provide information related to the identity of compounds present in an experimental (environmental) sample. The major advantage of bioassays is the possibility to investigate individual toxic effects very sensitively and specifically and in a quantitative manner. Therefore they are preferred for the monitoring and characterisation of individual toxic effects. It is virtually impossible to identify compounds with genomics tools. Their major advantages are the possibilities to investigate multiple toxicological endpoints in a (semi)qualitative manner.

Table 2: Summary of the expected possibilities and impossibilities of current water quality monitoring approaches (chemical analysis and *in vitro* bioassays) and genomics tools as discussed in this chapter. A multifold of plusses indicates a stronger relevance. Question marks indicate that confirmation of the expectation in practice is still needed.

Target	Monitoring compounds + unknowns	Monitoring effects	Characterisation multiple toxic effects	Characterisation individual toxic effects
Analytical chemistry	++++	-	-	-
Bioassay	=	++++	-	++++
Genomics: indiv compounds	-	+++?	++++	+++
Genomics environmental samples / mixtures	-	++?	+++?	++?

In conclusion, the following potential applications for genomics in water quality assessment are foreseen:

- Prediction of the toxicity of compounds that are newly detected in aquatic samples and of which possible toxic properties and MOA are unknown.
- Environmental monitoring of toxic pressure at locations relevant for the water sector, such as drinking water intake locations. For this purpose, the NOTEL concept is a helpful means to mutually compare different sites.
- Pre-screening of chemicals or locations for relevant toxicological effects, that can be investigated in more depth with specific *in vitro* bioassays

At present, no standardized genomics-based risk assessment tools are available for the assessment of water quality. However, 'omic' techniques have the potential to become very powerful methods in this field. It is therefore strongly advised that developments in this field should be followed intensively.

Increasing the use of 'omic' methods in chemical risk assessment and environmental monitoring requires an expanded toxicogenomics reference database and a better understanding of the relationships between specific responses and biomarkers to adverse effects. Furthermore, it is recommended to establish consortia for (eco)toxicogenomics studies to overcome the challenges as mentioned earlier in this report (Halligan *et. al.* 2008, Poynton *et. al.*, 2008). The major challenge will be the interpretation of data when

cells or organisms are exposed to complex mixtures of chemical substances, as present in the drinking water sources. Moreover, the impact of confounding factors should be separated from toxic responses.

The involvement of the Joint Research Programme of the Dutch drinking water Companies (BTO), which is funding the present report, in a Dutch consortium that focuses on the potential applications of 'omics' for assessing water quality is an important first step. This consortium should follow state of the art developments in the field and should design and perform pilot studies to demonstrate the feasibility of genomics techniques for water quality assessment. The ultimate goal of the research should be to design a monitoring tool that watches over the quality of the water cycle. This tool should provide an early-warning for potential chronic effects of micropollutants in our drinking water sources.

8 Acknowledgements

The authors wish to thank dr. Tim Williams, prof. dr. Kevin Chipman (School of Biosciences, The University of Birmingham, United Kingdom), Sander van der Linden and dr. Bart van der Burg (BioDetection Systems B.V., The Netherlands) for provision of helpful comments on the report.

9 References

Aebersold, R. and Mann, M. (2003). Mass spectrometry-based proteomics. Nature 422,198–207.

Ankley, G., Miracle, A., Perkins, E., Daston, G. (2008). Genomics in Regulatory Ecotoxicology. Pensacola, FL:SETAC Press.

Bogers, R., de Vries-Buitenweg, S., Geuijen, L., van der Waart, B., Kuiper, R., van der Linden, S., Puijker, L., Murk, A., van der Burg, B., Legler, J. (2007). An *in vitro/in vivo* screening assay as a sensitive tool to assess endocrine disruptive activity in surface water. Environ Int 33(3), 292-301.

Boverhof, D.R., Zacharewski, T.R. (2006). Toxicogenomics in risk assessment: applications and needs. Toxicol Sci 89(2), 352–360.

Botsford, J.L. (2002). A comparison of ecotoxicological tests. Altern Lab Anim 30, 539–550.

Brown, S.C., Kruppa, G. and Dasseux, J.L. (2005). Metabolomics applications of FT-ICR mass spectrometry Mass Spectrom Rev 24, 223–231.

Bultelle, F., Panchout, M., Leboulenger, F. and Danger, J.M. (2002). Identification of differentially expressed genes in *Dreissena polymorpha* exposed to contaminants. Mar Environ Res 54, 385–389.

Burczynski, M.E. (Eds) (2003). An introduction to toxicogenomics; CRC Press, Boca Raton (Florida, USA).

Chang, L. W., Toth, G.P., Gordon, D.A., Graham, D.W., Meier, J.R., Knapp, C.W., Denoyelles, F.J., Jr., Campbell, S. and Lattier, D.L. (2005). Responses of molecular indicators of exposure in mesocosms: Common carp (*Cyprinus carpio*) exposed to the herbicides alachlor and atrazine. Environ Toxicol Chem 24, 190–197.

De Longueville, F., Bertholet, V., Remacle, J. (2004). DNA microarrays as a tool in toxicogenomics, Comb Chem High Throughput Screen 7, 207-11.

Denslow N.D, Kocerha, J., Sepulveda, M.S., Gross, T., Holm, S.E. (2004). Gene Expression Fingerprints of Largemouth Bass (*Micropterus salmoides)* Exposed to Pulp and Paper Mill Effluents. Mutation Res 552, 19-34.

Diatchenko, L., Lau, Y.F., Campbell, A.P., Chenchik, A., Moqadam, F., Huang, B., Lukyanov, S., Lukyanov, K., Gurskaya, N., Sverdlov, E.D. and Siebert, P.D. (1996). Suppression subtractive hybridization: A method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci USA 93, 6025–6030.

Ellinger-Ziegelbauer, H., Stuart, B., Wahle, B., Bomann, W., Ahr, H.J. (2004). Characteristic expression profiles induced by genotoxic carcinogens in rat liver. Toxicol Sci 77, 19-34.

Falciani, F., Diab, A.M., Sabine, V., Williams, T.D., Ortega, F., George, S.G., Chipman, J.K. (2008). Hepatic transcriptomic profiles of European flounder (*Platichthys flesus*) from field sites and computational approaches to predict site from stress gene responses following exposure to model toxicants. Aquat Toxicol 90, 92-101.

Filby, A.L., Thorpe, K.L., Maack, G., Tyler, C.R. (2007). Gene Expression Profiles Revealing the Mechanisms of Anti-Androgen-and Estrogen-Induced Feminization in Fish. Aquat Toxicol 81, 219-231.

Forbes, V.E., Palmqvist, A., Bach, L., 2006. The use and misuse of biomarkers in ecotoxicology. Environ Toxicol Chem 24, 272-280.

Freeman, K. (2004). Toxicogenomics data: the road to acceptance, Environ Health Perspect 112, A678-85.

Gevaert, K. and Vandekerckhove, J. (2000). Protein identification methods in proteomics. Electroph 21, 1145–1154.

Gowda, M., Li, H., Alessi, J., Chen, F., Pratt, R., Wang, G.-L. (2006). Robust Analysis of 5'-Transcript Ends (5'-RATE): A Novel Technique for Transcriptome Analysis and Genome Annotation. Nucl Acids Res 34(19), e126.

Guerreiro, N., Staedtler, F., Grenet, O., Kehren, J., Chibout, S.D. (2003). Toxicogenomics in drug development. Toxicol Pathol 31, 471-479.

Guilhermino, L., Diamantino, T., Silva, M.C. and Soares, A.M. (2000). Acute toxicity test with *Daphnia magna*: An alternative to mammals in the prescreening of chemical toxicity? Ecotoxicol Environ Saf 46, 357–362.

Halligan, E.P. and Lunec, J. (2008). Toxicogenomics: unlocking the potential of the human genome. In Comparative toxicogenomics; Hogstrand, C. and Kille, P., Eds; Elsevier Science: Amsterdam, The Netherlands.

Hamadeh, H.K., Bushel, P.R., Jayadev, S., DiSorbo, O., Bennett, L., Li, L., Tennant, R., Stoll, R., Barrett, J. C., Paules, R.S., Blanchard, K., Afshari, C.A. (2002a). Prediction of compound signature using high density expression profiling. Toxicol Sci 67, 232–240.

Hamadeh, H.K., Bushel, P.R., Jayadev, S., Martin, K., DiSorbo, O., Siebert, S., Bennett, L., Tennant, R., Stoll, R., Barrett, J.C., Blanchard, K., Paules, R.S., Blanchard, K., and Afshari, C.A. (2002b). Gene expression analysis reveals chemical-specific profiles. Toxicol Sci 67, 219–231.

Hill, A.J., Teraoka, H., Heideman, W. and Peterson, R.E. (2005). Zebrafish as a model vertebrate for investigating chemical toxicity. Toxicol Sci 86, 6–19.

HGSC (Human Genome Sequencing Consortium), 2004. Finishing the euchromatic sequence of the human genome. Nature, 431, 931-45.

Houtman, C.J. (2010). Emerging contaminants in surface waters and their relevance for the production of drinking water in Europe, Journal of Integrative Environmental Sciences 7 (4), 271-295.

Kim, H.J., Ishidou, E., Kitagawa, E., Momose, Y., Iwahashi, H. (2004). A yeast DNA microarray for the evaluation of toxicity in environmental water containing burned ash, Environ Monit Assess 92, 253-72.

Kim, J.Y., Kwon, J., Kim, J.E., Koh, W.S., Chung, M.K., Yoon, S., Song, C.W., Lee, M. (2005). Identification of potential biomarkers of genotoxicity and carcinogenicity in L5178Y mouse lymphoma cells by cDNA microarray analysis, Environ Mol Mutagen 45, 80-89.

Krasnov, A., Afanasyev, S., Oikari, A. (2007). Hepatic Responses of Gene Expression in Juvenile Brown Trout (*Salmo trutta lacustris*) Exposed to Three Model Contaminants Applied Singly and in Combination. Environ Toxicol Chem 26(1),100-109.

Larkin, P., Knoebl, I. Denslow, N.D. (2003a). Differential gene expression analysis in fish exposed to endocrine disrupting compounds. Comp Biochem Physiol B 136B, 149–161.

Larkin, P., Sabo-Attwood, T., Kelso, J. Denslow, N. D. (2003b). Analysis of gene expression profiles in largemouth bass exposed to 17-beta-estradiol and to anthropogenic contaminants that behave as estrogens. Ecotoxicol 12, 463–468.

Larkin, P., Villeneuve, D.L., Knoebl, I., Miracle, A.L., Carter, B.J., Liu, L., Denslow, N.D., Ankley, G.T. (2007). Development and Validation of a 2,000-Gene Microarray for the Fathead Minnow (*Pimephales promelas*). Environ Toxicol Chem 26(7), 1497-1506.

Liang, P. and Pardee, A.B. (1992). Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. Science 257, 967–971.

Lobenhofer, E.K., Cui, X., Bennett, L., Cable, P.L., Merrick, B.A., Churchill, G.A. and Afshari, C.A. (2004). Exploration of low-dose estrogen effects: Identification of No Observed Transcriptional Effect Level (NOTEL). Toxicol Pathol 32, 482–492.

Lockhart, D.J., Dong, H., Byrne, M.C., Follettie, M.T., Gallo, M.V., Chee, M.S., Mittmann, M., Wang, C., Kobayashi, M., Horton, H. and Brown, E.L. (1996). Expression monitoring by hybridization to high-density oligonucleotide arrays. Nat Biotechnol. 14, 1675–1680.

Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.-J., Chen, Z., Dewell, S.B., Du, L., Fierro, J.M., Gomes, X.V., Godwin, B.C., He, W., Helgesen, S., Ho, C.H., Irzyk, G.P., Jando, S.C., Alenquer, M.L.I., Jarvie, T.P., Jirage, K.B., Kim, J.-B., Knight, J.R., Lanza, J.R., Leamon, J.H., Lefkowitz, S.M., Lei, M., Li, J., Lohman, K.L., Lu, H., Makhijani, V.B., McDade, K.E., McKenna, M.P., Myers, E.W., Nickerson, E., Nobile, J.R., Plant, R., Puc, B.P., Ronan, M.T., Roth, G.T., Sarkis, G.J., Simons, J.F., Simpson, J.W., Srinivasan, M., Tartaro, K.R., Tomasz, A., Vogt, K.A., Volkmer, G.A., Wang, S.H., Wang, Y., Weiner, M.P., Yu, P., Begley, R.F., Rothberg, J.M. (2005). Genome Sequencing in Microfabricated High- Density Picolitre Reactors. Nature 437(7057), 376-380.

Mattingly, C.J., Rosenstein, M.C., Davis, A.P., Colby, G.T., Forrest, J.N., Jr. Boyer, J.L. (2006). The comparative toxicogenomics database: A cross-species resource for building chemical-gene interaction networks. Toxicol Sci 92, 587–595.

Medlin, J.F. Timely toxicology, Environ Health Perspect, 1999 (107) A256-8, 16.

Meyer, J.N., Volz, D.C., Freedman, J.H., Di Giulio, R.T. (2005). Differential Display of Hepatic MRNA From Killifish (*Fundulus heteroclitus*) Inhabiting a Superfund Estuary. Aquat Toxicol 73(4), 327-341.

Moggs, J.G., 2005. Molecular Responses to Xenoestrogens: Mechanistic Insights From Toxicogenomics. Toxicology 213(3), 177-193.

Moens, L.N., van der Ven, K., Van Remortel, P., Del-Favero, J., De Coen, W.M. (2006). Expression Profiling of Endocrine-Disrupting Compounds Using a Customized *Cyprinus carpio* cDNA Microarray. Toxicol Sci 93(2), 298-310.

Muellner, M.G., Attene-Ramos, M.S., Hudson, M.E., Wagner, E.D., Plewa, M.J. (2010). Human cell toxicogenomic analysis of bromoacetic acid: a regulated drinking water disinfection by-product. Environ Mol Mutagen 51, 205-214.

Newton, R.K., Aardema, M., Aubrecht, J. (2004). The utility of DNA microarrays for characterizing genotoxicity. Environ Health Perspect 112, 420-2.

Nicholson, J.K., Lindon, J.C., Holmes, E. (1999). 'Metabonomics': Understanding the metabolic responses of living systems to pathophysiological stimuli via multivariate statistical analysis of biological NMR spectroscopic data. Xenobiotica 29, 1181–1189.

NRC (National Research Council). 2007. Toxicity Testing in the Twenty-first Century: A Vision and a Strategy. Washington, DC:National Academies Press.

Nuwaysir, E.F., Bittner, M., Trent, J., Barrett, J.C. and Afshari, C.A. (1999). Microarrays and toxicology: The advent of toxicogenomics. Mol Carcinogen 24, 153–159.

Paules, R. (2003). Phenotypic anchoring: Linking cause and effect. Environ Health Perspect 111, A338–A339.

Perkins E.J, Chipman, J.K., Edwards, S., Habib, T., Falciani, F., Taylor, R., Van Aggelen, G.,Vulpe, V., Antczak, P., Loguinov, A. (2011). Reverse Engineering Adverse Outcome Pathways. Environ Toxicol Chem 30(1), 22-38.

Plumb, R.S., Stumpf, C.L., Gorenstein, M.V., Castro-Perez, J.M., Dear, G.J., Anthony, M., Sweatman, B.C. and Haselden, J.N. (2002). Metabonomics: The use of electrospray mass spectrometry coupled to reversed-phase liquid chromatography shows potential for the screening of rat urine in drug development. Rapid Commun Mass Spectrom 16, 1991–1996.

Poynton, H.C., Varshavsky, J.R., Chang, B., Cavigiolio, G., Chan, S., Holman, P.S., Loguinov, A.V., Bauer, D.J., Komachi, K., Theil, J.E., Perkins, E.J., Hughes, O., Vulpe, C.D. (2007). *Daphnia Magna* Ecotoxicogenomics Provides Mechanistic Insights Into Metal Toxicity. Environ Sci Technol 41(3), 1044-1050.

Poynton, C.H., Wintz, H., Vulpe, C.D. (2008). Progress in ecotoxicogenomics for environmental monitoring, mode of action and toxicant identification. In Comparative toxicogenomics, Hogstrand, C. and Kille, P., Eds, Elsevier Science: Amsterdam, The Netherlands.

Regulation (EC) No 1907/2006 of the European Parliament and of the Council of 18 December 2006 concerning the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH), establishing a European Chemicals Agency.

Roling, J.A., Bain, L.J., Gardea-torresdey, J., Key, P.B., Baldwin, W.S. (2007). Using Mummichog (*Fundulus heteroclitus*) Arrays to Monitor the Effectiveness of Remediation at a Superfund Site in Charleston, South Carolina, USA. Environ Toxicol Chem 26(6), 1205-1213.

Schena, M., Shalon, D., Davis, R.W., Brown, P.O. (1995). Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 270, 467–470.

Schriks, M., Heringa, M.B., van der Kooi, M.E., de Voogt, P., van Wezel, A.P. (2010). Toxicological relevance of emerging contaminants for drinking water quality. Water res 44, 461-476.

Shepard, J.L. and Bradley, B.P (2000). Protein expression signatures and lyosomal stability in *Mytilus edulis* exposed to graded copper concentrations. Mar Environ Res 50, 457-463.

Shioda, T. (2004). Application of DNA microarray to toxicological research, J Environ Pathol Toxicol Oncol 23, 13-31.

Snape, J.R., Maund, S.J., Pickford, D.B. and Hutchinson, T.H. (2004). Ecotoxicogenomics: The challenge of integrating genomics into aquatic and terrestrial ecotoxicology. Aquat Toxicol 67, 143–154.

Snell, T.W., Brogdon, S.E., Morgan, M.B. (2003). Gene expression profiling in ecotoxicology. Ecotoxicology 12, 475–483.

Soetaert, A., Moens, L.N., Van der Ven, K., Van Leemput, K., Naudts, B. Blust, R., De Coen, W.M. (2006). Molecular Impact of Propiconazole on *Daphnia Magna* Using a Reproduction-Related CDNA Array. Comp Biochem Physiol C: Comp Pharmacol Toxicol 142(1-2), 66-76.

Subramanian, A., Tamayoa, P., Moothaa, V.K., Mukherjeed, S., Eberta, B.L., Gillettea, M.A., Paulovichg, A., Pomeroyh, S.L., Golub, T.R., Lander, E.S., Mesirova, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. PNAS 102, 15545-15550.

Sumpter, J.P. and Jobling, S. (1995). Vitellogenesis as a biomarker for estrogenic contamination of the aquatic environment. Environ Health Perspect 103 (Suppl. 7), 173–178.

Steinberg C.E.W., Stürzenbaum, S.R., Menzel, R. (2008). Genes and environment — Striking the fine balance between sophisticated biomonitoring and true functional environmental genomics. Sci Total Environ 400, 142-161

Unlu, M., Morgan, M.E. and Minden, J.S. (1997). Difference gel electrophoresis: A single gel method for detecting changes in protein extracts. Electroph 18, 2071–2077.

Schwarzenbach, R.P., Escher, B.I., Fenner, K., Hofstetter, T.B., Johnson, C.A., von Gunten, U., Wehrli, B., 2006. The challenge of micropollutants in aquatic systems. Science 313, 1072-1077.

US EPA 2002. Interim Policy on Genomics: Science Policy Council. Washington, DC:U.S. Environmental Protection Agency)

U.S. EPA (Environmental Protection Agency). (2004). Potential implications of genomics for regulatory and risk assessment applications at EPA. Genomics Task Force Workgroup, Science Policy Council, Washington, DC, EPA 100/B-04/002. Available online at http://www.epa.gov/OSA/pdfs/EPA-Genomics-White-Paper.pdf.

Van Aggelen, G., Ankley, G.T., Baldwin, W.S., Bearden, D.W., Benson, W.H., Falciani, F., George, S.G., Helbing, C.C., Hoekstra, P.F., Iguchi, T., Kagami, Y., Katsiadaki, I., Kille, P., Liu, L., Lord, P.G., McIntyre, T., O'Neil, A., Osashoff, H., Perkins, E.J., Santos, E.M., Skirrow, R.C., Snape, J.R., Tyler, C.R., Versteeg, B., Viant, M.R., Volz, D.C., Williams, T.D., Yu, L. (2010). Integrating omic technologies into aquatic ecological risk assessment and environmental monitoring: hurdles, achievements, and future outlook. Environ Health Perspect 118, 1-5.

Van Boxtel, A.L., Kamstra, J.H., Cenijn, P.H., Pieterse, B., Wagner, J.M., Antink, M., Krab, K., van der Burg B., Marsh, G., Brouwer, A., Legler, J. (2008). Microarray analysis reveals a mechanism of phenolic polybrominated diphenylether toxicity in zebrafish. Environ Sci Technol 42(5), 1773-1779.

Van Delft, J. H., van Agen, E., van Breda, S.G., Herwijnen, M.H., Staal, Y.C., Kleinjans, J.C. (2004). Discrimination of genotoxic from non-genotoxic carcinogens by gene expression profiling. Carcinog 25, 1265-1276.

Van der Oost, R., J. Beyer and N.P.E. Vermeulen, 2003. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. Environ. Toxicol. Pharmacol. 13: 57-149.

Vrana, K.E., Freeman, W.M., Aschner, M. (2003). Use of microarray technologies in toxicology research, Neurotoxicol 24, 321-332.

Walker, C.H., Hopkin, S.P., Sibly, R.M., Peakall, D.B., 2006. Principles of Ecotoxicology. Taylor and Francis Group, Boca Raton, Florida.

Williams, T.D., Gensberg, K., Minchin, S.D., Chipman, J.K. (2003). A DNA expression array to detect toxic stress response in European flounder (*Platichthys flesus*). Aquatic Toxicol 65, 143-157.

Wintz, H., Yoo, J.L., Loguinov, A., Wu, Y.Y., Steevens, J.A., Holland, R.D., Beger, R.D., Perkins, E.J., Hughes, O. and Vulpe, C.D. (2006). Gene expression profiles in fathead minnow exposed to 2,4-DNT: Correlation with toxicity in mammals. Toxicol Sci 94, 71–82.

Whyte, J.J., Jung, R.E., Schmitt, C.J., Tillitt, D.E. (2000). Ethoxyresorufin-O-deethylase (EROD) activity in fish as a biomarker of chemical exposure. Crit Rev Toxicol 30, 347-570.

Yamamoto, M., Wakatsuki, T., Hada, A. and Ryo, A. (2001). Use of serial analysis of gene expression (SAGE) technology. J Immunol Methods 250, 45–66.

Postbus 1072 3430 BB Nieuwegein The Netherlands P +31 30 606 95 11 F +31 30 606 11 65 E info@kwrwater.nl